General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

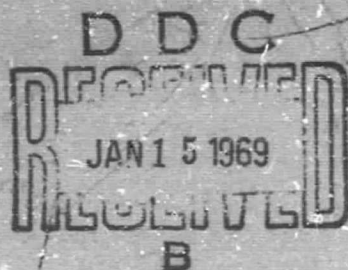Produced by the NASA Center for Aerospace Information (CASI)

THEMIS SIGNAL ANALYSIS STATISTICS RESEARCH PROGRAM

ASYMPTOTIC INDEPENDENCE BETWEEN LARGEST AND SMALLEST

OF A SET OF INDEPENDENT OBSERVATIONS

by

John E. Walsh

Technical Report No. 17
Department of Statistics THEMIS Contract

Department of Statistics
Southern Methodist University
Dallas, Texas   75222

THEMIS SIGNAL ANALYSIS STATISTICS RESEARCH PROGRAM

ASYMPTOTIC INDEPENDENCE BETWEEN LARGEST AND SMALLEST

OF A SET OF INDEPENDENT OBSERVATIONS

by

John E. Walsh

Technical Report No. 17
Department of Statistics THEMIS Contract

October 4, 1968

DEPARTMENT OF STATISTICS
Southern Methodist University

ASYMPTOTIC INDEPENDENCE BETWEEN LARGEST AND SMALLEST

OF A SET OF INDEPENDENT OBSERVATIONS

John E. Walsh

Southern Methodist University*

ABSTRACT

Let X_n and X_1 be the largest and smallest order statistics,
respectively, of a set of n independent univariate observations.
Under rather general conditions, with respect to the distributions
of the individual observations, X_n and X_1 are asymptotically
independent. That is, the maximum difference between $P(X_1 \leq x_1, X_n \leq x_n)$
and $P(X_1 \leq x_1)P(X_n \leq x_n)$ tends to zero as $n \rightarrow \infty$. However, asymptotic
independence does not occur for all cases.

INTRODUCTION AND RESULTS

Asymptotic independence of the largest and smallest order
statistics of a random univariate sample is well known. The question
arises as to what extent this asymptotic independence remains
when the observations are still required to be independent but can
have arbitrarily different distributions. That is, let X_n and X_1
be the largest and smallest, respectively of a set of n independent
observations. When does the maximum of

$$P(X_1 \le x_1)P(X_n \le x_n) - P(X_1 \le x_1, X_n \le x_n), \tag{1}$$

over x_1 and x_n, tend to zero as $n \to \infty$?

The interest is in the range of x_1 values that are meaningful for $P(X_1 \le x_1)$ and range of x_n values that are meaningful for $P(X_n \le x_n)$. That is, the analysis is made in terms of (attainable) percentiles for $P(X_1 \le x_1)$ and for $P(X_n \le x_n)$. Let

$$P(X_n \le x_n) = e^{-a}, \quad P(X_1 \le x_1) = 1 - e^{-b},$$

where $\underline{a}$ and $\underline{b}$ are arbitrary but fixed. Then,

$$\prod_{i=1}^{n} F_i(x_n) = e^{-a}, \quad \prod_{i=1}^{n} [1 - F_i(x_1)] = e^{-b},$$

where $F_i(x)$ is the cumulative distribution function (cdf) for the i-th observation $(i = 1, \ldots, n)$.

Now consider the $F_i(x_n)$ and $F_i(x_1)$. Let these cdf's be expressed as

$$F_i(x_n) = e^{-a_i/n}, \quad F_i(x_1) = 1 - e^{-b_i/n}$$

where $a_i = a_i(n)$ and $b_i = b_i(n)$. Asymptotic independence between X_n and X_1 always occurs if

$$a_i \le A(n), \quad b_i \le B(n)$$

for all i. Here, $A(n)$ and $B(n)$ are $O(n)$ and at least one of them is $o(n)$. That is both $A(n)/n$ and $B(n)/n$ tend to constants as $n \to \infty$, and at least one of these constants is zero. For example, the forms $C_1 n/\log n$ and $C_2 n^{1-\epsilon}$ (with $\epsilon > o$ and fixed but as small as desired) yield zero constants.

Examination shows that asymptotic independence fails to occur only when, for one or more values of i, both a_i and b_i are $O(n)$ and neither is $o(n)$. For these observations, the values of $1 - F_i(x_n)$ and $F_i(x_1)$, representing the "tail" probabilities, are relatively much larger than these values for the other observations (ratio becomes infinite as $n \to \infty$). Thus, for large n, approximate independence of X_n and X_1 should not be accepted when a few of the distributions seem to have a much wider spread (in both tails) than the others.

The next and final section contains derivations of the results that are stated in this section.

DERIVATIONS

The difference (1) can be written

$$\prod_{i=1}^{n} F_i(x_n)\left[1 - F_i(x_1)\right] - \prod_{i=1}^{n} \left[F_i(x_n) - F_i(x_1)\right]$$

$$= e^{-(a+b)} - e^{-a} \prod_{i=1}^{n}\left[1 - e^{a_i/n} + e^{-(b_i - a_i)/n}\right]$$

$$= e^{-(a+b)} - e^{-a} \prod_{i=1}^{n}\left[1 - \frac{b_i}{n} - \sum_{k=2}^{\infty} \frac{a_i^k + (-1)^{k+1}(b_i-a_i)^k}{k!\, n^k}\right]$$

$$= e^{-(a+b)} - e^{-a} \exp\left\{\sum_{i=1}^{n} \log_e\left[1 - \frac{b_i}{n} - \sum_{k=2}^{\infty} \frac{a_i^k + (-1)^{k+1}(b_i-a_i)^k}{k!\, n^k}\right]\right\}$$

$$= e^{-(a+b)}\left(1 - \exp\left\{-\sum_{i=1}^{n} \frac{a_i b_i}{n^2}\left[1 + \frac{a_i+b_i}{2n} + \sum_{k=4}^{\infty} G_{k-2}\left(\frac{a_i}{n}, \frac{b_i}{n}\right)\right]\right\}\right),$$

where $G_{k-2}(a_i/n, b_i/n)$ is a mixed polynomial of degree $k - 2$ that is symmetrical in a_i/n and b_i/n. The value of

$$\sum_{i=1}^{n} \frac{a_i b_i}{n^2} \left[1 + \frac{a_i + b_i}{2n} + \sum_{k=4}^{\infty} G_{k-2}\left(a_i/n, b_i/n\right) \right] \qquad (2)$$

is largest when some of the a_i have their maximum value, some of the b_i have their maximum value, and the others are zero. Also, the i values are such that the summation of $a_i b_i / n^2$ is largest and, say, $A(n)$ is $O(n)$ with nonzero constant (equivalent results would be obtained if $B(n)$ had the nonzero constant).

Let the values of i such that $a_i = A(n)$ be $i = 1, \ldots, r_n$, while the values such that $b_i = B(n)$ are $i = 1, \ldots, r_1$. Then, since,

$$\sum_{i=1}^{n} a_i = na, \qquad \sum_{i=1}^{n} b_i = nb,$$

r_n is $O(1)$ and r_1 is $O[n/B(n)]$. With these substitutions, (2) becomes

$$r_n [A(n)/n][B(n)/n][1 + O(1)]$$

and tends to zero as $n \to \infty$. Thus, the exponential of the negative of (2) tends to unity and the difference (1) tends to zero.

If both $A(n)$ and $B(n)$ had nonzero constants, the value of (2) would be

$$\min(r_1, r_n)[A(n)/n][B(n)/n][1 + O(1)]$$

and would not tend to zero as $n \to \infty$. Hence the difference (1) would not tend to zero and asymptotic independence does not occur.

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*	2a. REPORT SECURITY CLASSIFICATION
SOUTHERN METHODIST UNIVERSITY	UNCLASSIFIED
	2b. GROUP
	UNCLASSIFIED

3. REPORT TITLE

Asymptotic Independence Between Largest and Smallest
of a Set of Independent Observations

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Technical Report

5. AUTHOR(S) *(First name, middle initial, last name)*

John E. Walsh

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
October 4, 1968	4	0
8a. CONTRACT OR GRANT NO. N00014-68-A-0515	9a. ORIGINATOR'S REPORT NUMBER(S)	
b. PROJECT NO. NR 042-260	17.	
c.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*	
d.		

10. DISTRIBUTION STATEMENT

No limitations

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Office of Naval Research

13. ABSTRACT

Let X_n and X_1 be the largest and smallest order statistics,
respectively, of a set of n independent univariate observations.
Under rather general conditions, with respect to the distributions
of the individual observations, X_n and X_1 are asymptotically
independent. That is, the maximum difference between $P(X_1 \leq x_1, X_n \leq x_n)$
and $P(X_1 \leq x_1)P(X_n \leq x_n)$ tends to zero as $n \to \infty$. However, asymptotic
independence does not occur for all cases.